

**When are our forecasts ahead better, i.e., more accurate?**

Konner Macias  
004603916

Variable Name	R Name	Description and Source	Training Set	Test Set
HOUSTNSA	hs	Housing Starts: Total new privately owned housing units (in thousands). Monthly. Not seasonally adjusted <b>This is the variable we want to forecast</b>	Jan 1, 1959 Aug 1, 2017	Sept 1, 2019 Aug 1, 2018
LNU04000002	uw	Unemployment rate: Women (percent) Monthly. Not seasonally adjusted	Jan 1, 1959 Aug 1, 2017	
UNRATNSA	ur	Civilian Unemployment Rate (percent) Monthly. Not seasonally adjusted	Jan 1, 1959 Aug 1, 2017	

## 1 Introduction

The main question we are trying to address in this paper is: **When are our forecasts ahead better, i.e., more accurate?**

- (a) When the only information we use is the historical information of the time series we want to forecast, and nothing else, as when using SARIMA(p,d,q)(P,D,Q) or perhaps SARIMA with garch?
- (b) When the variable we want to forecast is the dependent variable in a traditional regression model and we use other information in the form of other variables that play the role of exogenous variables? Such model could also have dummies for seasonals and polynomial trends for trend. The residuals of such a model probably need to be modeled as AR or ARIMA model to account for autocorrelation. This is the approach we used when we studied gls to approach regression with autocorrelated residuals.
- (c) When we use information in other variables, but the variable we want to forecast is a dependent and independent variable at the same time, and so are the other variables both dependent and independent, as in VAR?
- (d) When we average the forecasted values of (a), (b), (c), to obtain a consensus forecast.

These questions always arise in practice in a number of sciences, among them economics and meteorology. Nobody would ever use a single model to make a forecast for time  $t + 1$ . Almost all areas that predict the future fit several models, and then they average the forecasts obtained for each future time  $t + 1$  from all the models to obtain what they called the consensus forecast for time  $t$ . Google the word consensus forecast to learn more. The Blue Chip consensus forecasts of economic indicators are an example of how the many forecasting companies out there combine their forecasts into a consensus forecast.

## 2 Data

In this section, we will describe that data that we will use to answer the question posed in the introduction. The variable that we will want to forecast is housing starts in the United States. Housing starts is considered to be a leading indicator of what might come next in the economy. If housing construction starts to flourish that should be an indication of prosperity to come, some economists think. Other information that will be used for the models that use the other variables is unemployment rate, and unemployment rate for women. The source of the data is FRED (Federal Reserve Economic Data, <https://fred.stlouisfed.org>) The variables are observed from January 1st 1959 to August 1st 2018. We will use January 1st 1959 to August 1st 2017 as training data, to fit the model, and then we will forecast from September 1st 2017 to August 1st 2018. TODO All this is too many words, and a good paper would summarize all the information regarding the data, training set and test set in a table as follows: